A genetic analysis of bacterial glycogen branching enzymes

Kiel, Jan Andries Kornelis Willem

SUMMARY AND GENERAL DISCUSSION

Potato starch is widely used in the food industry. Modification of starch requires its solubilization in water. However, watery solutions of starch appear to be rather unstable at low temperatures, because the amylose portion of starch tends to crystallize in a process known as retrogradation. Since the amylopectin component of starch is not subject to retrogradation, stable starch solutions at low temperatures might be obtained by converting the amylose component to amylopectin. This can be effected by the action of branching enzymes, which under natural conditions introduce the α-1,6-linkages (branches) present in the storage glucans of plants, animals and many micro-organisms. However, the branching reaction would have to be carried out in watery starch solutions. This implies that the reaction temperature must be high in order to prevent retrogradation of the substrate. Therefore, a suitable branching enzyme should have the ability to introduce α-1,6-branches in starch solutions at high temperatures. The main purpose of the investigations described in this Thesis was to isolate and characterize genes encoding such thermostable branching enzymes.

In Chapter 1 a general introduction is given discussing glycogen biosynthesis in bacteria. The main conclusion from this literature survey is that the presence of glycogen in bacteria (and thus presumably also of branching enzyme) does not seem to be restricted to particular classes of bacteria. This indicates that bacteria that are able to grow at elevated temperatures, which reputedly have thermostable enzymes, may posses suitable branching enzymes. However, it has been reported that the mesophilic cyanobacterium *Anacystis nidulans* had a branching enzyme with an optimal temperature for activity between 40 and 60°C.

Chapter 2 describes the development of a technique by which mutations can be introduced into the *Escherichia coli* chromosome by gene replacement between the chromosome and a plasmid carrying the mutant gene. The segregational instability of plasmids in *E. coli* was used to isolate with high efficiency *E. coli* mutants. By this method we constructed an internally deleted *glgB* mutant of *E. coli*, which lacked the branching enzyme. This mutant was used in subsequent experiments to clone genes encoding heterologous branching enzymes.

In Chapter 3 a fragment of the *E. coli glgB* gene was used as a heterologous DNA probe to isolate the gene encoding the branching enzyme from the cyanobacterium *Synechococcus* sp. PCC7942 (also known as *Anacystis nidulans* R2). The gene appeared to be inactive in *E. coli*, either indicating that the promoter of the gene had not been cloned or that it was not recognized by *E. coli*. When preceded by a suitable

promoter the cloned DNA fragment specified a branching activity. The cloned gene specified two proteins of 84 kDa and 72 kDa in *E. coli*, respectively, which presumably resulted from translation-initiation at two different start codons. Furthermore, at least the shorter protein appeared to have branching activity.

In contrast to data from the literature, the cloned enzyme did not appear to be thermostable; the enzyme was optimally active at 35°C which is in accord with the mesophilic nature of *Synechococcus* sp.

In Chapter 4 the nucleotide sequence of the *Synechococcus* sp. *glgB* gene was determined. The sequenced region appeared to contain one complete open reading frame (ORF), which was followed by an incomplete ORF. The first ORF represented the coding region of the *Synechococcus* sp. branching enzyme and the incomplete ORF that of the N-terminus of a protein with high sequence similarity to human and rat uroporphyrinogen decarboxylase, an enzyme involved in heme and chlorophyll biosynthesis. These data suggest that in *Synechococcus* sp. the organization of genes involved in glycogen biosynthesis is different from that described for *E. coli*, because in *E. coli* the other genes involved in glycogen metabolism are located directly downstream from the *glgB* gene on the bacterial chromosome. The deduced primary sequence of the cloned branching enzyme had extensive similarity to that of the *E. coli* branching enzyme, most notably in the middle portion of the protein (62 % identity). The protein encoded by the ORF had an M_r of 89,206 and presumably represents the 84 kDa protein observed earlier (see above). The smaller 72 kDa protein, shown to be encoded by the same gene, was considerably shortened at its N-terminus. Translation of this smaller protein presumably initiates at an alternative start codon. Although no proof through amino acid sequencing is available, the nucleotide sequences surrounding an ATT and a TTG codon have characteristics that make it tempting to assume that translation-initiation of this protein starts at (one of) these codons.

In order to express the *Synechococcus* sp. *glgB* gene in *Bacillus subtilis*, the gene was provided with *B. subtilis* specific transcription and translation signals by means of a gene fusion with the *Bacillus licheniformis penP* gene. The fusion gene appeared to encode a protein in *B. subtilis* with a optimal temperature for activity similar to the wild-type enzyme. In a control no significant activity was observed. These data suggest that *B. subtilis* does not synthesize significant amounts of branching enzyme. Furthermore, because the gene fusion was constructed with the larger of the two proteins encoded by *Synechococcus* sp. *glgB*, it is likely that also this protein possessed branching activity.

Chapter 5 describes the cloning, characterization and expression of the *glgB* gene from the thermophilic sporeforming bacterium *Bacillus stearothermophilus*. The gene was isolated using the *E. coli glgB* gene as

a heterologous DNA probe. The enzyme encoded by the *B. stearothermophilus glgB* gene had a M_r of 74,787. A comparison of the primary sequence of the enzyme to those of the *E. coli* branching enzyme and the larger of the two proteins encoded by *Synechococcus* sp. *glgB* revealed considerable similarities. However, the *B. stearothermophilus* enzyme was appreciably shortened at its N-terminus, and in this respect resembled the shorter of the two proteins encoded by *Synechococcus* sp. *glgB*. The codon usage and the very low G+C content of the *B. stearothermophilus glgB* gene were strikingly different from those reported for other genes isolated from thermophilic micro-organisms. In an expression vector the gene directed the synthesis of a thermostable branching enzyme in *E. coli* and *B. subtilis*. The optimal temperature for activity was 53°C.

The *B. stearothermophilus glgB* gene was preceded by a sequence resembling the consensus sequence of promoters recognized by the *B. subtilis* RNA polymerase containing the sporulation sigma factor H. Because an in frame *glgB-lacZ* fusion gene was exclusively active in a wild-type *B. subtilis* strain in the stationary phase, and not at all in a *B. subtilis spoOH* mutant, lacking the sigma factor H, it was concluded that in *B. stearothermophilus* expression of the branching enzyme occurs at the end of the logarithmic phase of growth and prior to sporulation. This is in accord with data from the literature indicating that glycogen synthesis in *B. stearothermophilus* occurs at the transition to the stationary phase.

In Chapter 6 the cloning and characterization of the *glgB* gene from the caldo-active bacterium *Bacillus caldolyticus* is described. In this case the *B. stearothermophilus glgB* gene was used as a heterologous DNA probe to isolate the gene. Nucleotide sequence analysis of the cloned gene revealed a 1998 bp ORF encoding an enzyme with an M_r of 78,087, which had a very high amino acid sequence similarity with the *B. stearothermophilus* branching enzyme (overall 76 % identity). However, the *B. caldolyticus* enzyme was considerably extended at its C-terminus. In addition, the putative *B. caldolyticus glgB* gene resembled the *B. stearothermophilus glgB* gene in that it was preceded by a putative promoter sequence recognized by the *B. subtilis* RNA polymerase containing the sigma factor H, indicating that also in *B. caldolyticus* glycogen synthesis may occur at the end of the logarithmic growth phase. A major difference with *B. stearothermophilus* was that the *B. caldolyticus glgB* gene did not encode a thermostable branching enzyme. When cloned on an expression vector the gene from this extremely thermophilic sporeformer appeared to direct the synthesis of a thermolabile branching enzyme in *B. subtilis* with an optimal temperature for activity of 39°C. This was a rather surprising finding considering the fact that *B. caldolyticus* has an optimal temperature for growth of 72°C. The cause underlying the large difference in the optimal temperature for activity between the branching enzymes from the two thermophilic *Bacillus* spp. is as yet unknown.

Further analysis of the nucleotide sequence of the cloned *B. caldolyticus* DNA fragment revealed that the 5' end of a putative *glgC* gene, encoding ADP-glucose pyrophosphorylase (an enzyme also involved in glycogen biosynthesis) partly overlapped the 3' end of the *B. caldolyticus glgB* gene. This indicates that in *B. caldolyticus* the genes involved in glycogen biosynthesis may be clustered on the bacterial chromosome. Furthermore, the high sequence similarity observed between the N-terminus of the putative *B. caldolyticus* ADP-glucose pyrophosphorylase and those of the *E. coli* and *Salmonella typhimurium* enzymes indicates that these allosteric enzymes may have similar effectors.

Chapter 7 describes the isolation and characterization of a part of a putative glycogen (*glg*) operon of *Bacillus subtilis*, a potentially useful micro-organism for increased expression of heterologous genes. The *B. stearothermophilus glgB* gene was used as a heterologous probe to isolate this DNA region from a *B. subtilis* strain, which does not normally synthesize glycogen. Nucleotide sequence analysis of the cloned DNA fragment revealed that the putative *glg* operon was located directly downstream from the *trnB* region (encoding 21 different tRNAs), which maps at approximately 275° on the *B. subtilis* chromosome. A promoter recognized by the *B. subtilis* RNA polymerase containing the sporulation type of sigma factor E, which was identified by other authors and designated *bvx*, was located upstream of the operon. However, because the *B. subtilis* strain used in this study did not synthesize glycogen, the operon was apparantly inactive. Thus the *bvx* promoter may not be used in this *B. subtilis* strain.

The cloned portion of the putative *glg* operon contained three ORFs. ORF 1, which was contained in the region hybridizing to the *B. stearothermophilus glgB* probe, appeared to encode the *B. subtilis* branching enzyme (M_r = 73,674), because the sequence similarity between the deduced amino acid sequence of the protein and those of the branching enzymes isolated from the thermophilic *Bacillus* spp. was high (60 %). Furthermore, in an expression vector the putative *B. subtilis glgB* gene directed the synthesis of a branching enzyme with an optimal temperature for activity of 27°C, in accordance with the mesophilic nature of *B. subtilis*.

ORF 2 presumably encoded the *B. subtilis* ADP-glucose pyrophosphorylase, because the amino acid sequence similarity of the ORF with ADP-glucose pyrophosphorylases isolated from enteric bacteria was extensive. Furthermore, the sequence similarity with the N-terminus of the putative *B. caldolyticus* ADP-glucose pyrophosphorylase was extremely high (85 % identity). Therefore, ORF 2 was identified as the *B. subtilis glgC* gene.

The product of the incomplete ORF 3 also showed significant amino acid sequence similarity to bacterial ADP-glucose pyrophosphorylases. Although the function of this protein is unknown, the location of the

ORF in the putative *glg* operon suggests that the protein is involved in glycogen metabolism. Therefore, ORF 3 was designated *glgD*.

Despite the fact that the *B. subtilis* strain used in this study did not synthesize glycogen, we were able to induce glycogen synthesis in this strain by the introduction of a *B. subtilis* specific promoter upstream of the putative *glg* operon on the bacterial chromosome. This suggests that the *B. subtilis glg* operon contained intact genes, which are presumably inactive in the *B. subtilis* strain used in this study, despite the presence of an intact *bvx* promoter. Furthermore, these data suggest that possibly also the *glgA* and *glgP* genes, encoding glycogen synthase and glycogen phosphorylase, are located downstream from *glgD*. This, however, remains to be established.

In conclusion, we have isolated the *glgB* genes from four different micro-organisms. Three of the branching enzymes encoded by these genes were expected to be thermostable. However, the data presented in this Thesis show that, with the exception of the *B. stearothermophilus* enzyme, all enzymes were rather thermolabile. The most surprising finding was that *B. caldolyticus* contained a thermolabile branching enzyme. To our knowledge it has never been reported that such an extremely thermophilic micro-organism possesses a thermolabile intracellular enzyme.

The fact that the *B. stearothermophilus glgB* gene encoded a thermostable branching enzyme makes this gene the most promising gene for large scale production of branching enzyme. To this purpose increased expression of the gene and the introduction of the gene in a suitable host is required. This host may be a *B. subtilis* mutant, which lacks the *glgB* gene, thus preventing recombination between the *B. stearothermophilus glgB* gene and the *B. subtilis* gene. Owing to the availability of the cloned *B. subtilis glgB* gene, this host can be easily constructed.